Deregulation and the Theory of Contestable Markets

Elizabeth E. Bailey†
William J. Baumol††

Until recently, the need to regulate monopoly was considered virtually axiomatic, and the imposition of rules governing entry, exit and pricing was deemed a priority. The deregulation movement has raised pragmatic questions about these orthodoxies, and, more recently, a new body of economic analysis called the theory of contestable markets has provided a conceptual basis for the view that many markets that are subject to economies of scale should not be regulated by the conventional methods.[1]

The new theory has shown that neither large size nor fewness of firms *necessarily* means that markets need function unsatisfactorily. Impediments to entry and exit, not concentration or scale of operations, may be the primary source of interference with the workings of the invisible hand. Indeed, because regulators have been predisposed to interfere with both entry and exit,[2] the new analysis suggests that they have been among the primary causes of unsatisfactory industry performance.

Contestability theory focuses increased attention upon entry barriers and redefines their character. Economies of scale, for example, have frequently been considered an impediment to entry; contestability analysis shows, however, that they need not permit excessive profits or prices or any of the other manifestations usually associated with market power, even when scale economies make an industry a natural monopoly or an

† Dean, Graduate School of Industrial Administration, Carnegie-Mellon University.
†† Professor of Economics, Princeton and New York Universities. The authors are grateful to George Eads, Ann Friedlaender, Judy Gelman, Daniel Kaplan, Steven Salop and David Sibley for their extremely valuable comments on an earlier version of this paper.

1. The tradition which asserts that even in oligopolistic industries barriers to entry are not always sufficiently great to permit monopoly profits traces its roots to Chadwick, *Results of Different Principles of Legislation and Administration in Europe: of Competition for the Field as Compared with Competition Within the Field of Service*, 22 J. ROYAL STATISTICAL SOC'Y 381 (1859). A similar position is taken in Demsetz, *Why Regulate Utilities?*, 11 J. L. & ECON. 55 (1968). These conclusions have recently been extended to the multiproduct case, and derived from a formal theory applicable to the full spectrum of market forms. *See* W. BAUMOL, J. PANZAR & R. WILLIG, CONTESTABLE MARKETS AND THE THEORY OF INDUSTRY STRUCTURE (1982). Much of the material in the first half of this paper draws on this recent work. For a measured appraisal, see Spence, *Contestable Markets and the Theory of Industry Structure*, 21 J. ECON. LITERATURE 981 (1983).

2. There is no shortage of examples. Thus, the ICC has long prevented railroads from abandoning unprofitable routes, while before deregulation, the CAB kept many potential entrants out of the most profitable markets.

oligopoly. It is the presence of sunk costs rather than economies of scale that is of vital significance for both theory and practice.

In sum, the theory of contestability calls for a major reorientation of both the charters and the operating programs of regulatory authorities. Many of its conclusions are consistent with the philosophy of deregulation. But the new analysis does not adopt the Panglossian view that a completely unconstrained free market is necessarily the best of all possible worlds. It claims only that where public measures are called for, the types of market intervention that ought to be undertaken are, in many cases, rather different from those that have traditionally been employed, and that there are *some* cases in which intervention is inappropriate even though it was previously thought to be desirable.

In the following pages, we begin by describing the basic principles of contestability theory. Next, we examine the efficiency attributes of contestable markets, the conditions that contribute to the contestability of a market, and the general implications for regulatory policy that follow from this analysis. Finally, we review some of the major regulatory reform activities that have been undertaken in Congress, the executive branch, and the independent regulatory agencies during the past few years and evaluate these moves in terms of contestability. We shall see that, by and large, the new directions of public policy have been remarkably consistent with what contestability theory would suggest.

I. Contestability Theory as an Alternative Ideal

A. *Perfect Contestability Contrasted With Perfect Competition*

The theoretical foundation of both regulatory and antitrust activity has traditionally relied heavily on the economic concept of perfect competition. Perfect competition has long been used as a standard ideal for the structure and performance of a market, though it is widely recognized to be unattainable in reality. The new analysis proposes the use of a different ideal: what is termed the "perfectly contestable market." Like perfect competition, this exemplary market form is undoubtedly unattainable except as an approximation. But while perfect competition and perfect contestability are both unattainable ideal states, there are many industries whose structure and performance may usefully be measured against the latter but not the former. To see why this is so let us first define the two concepts and indicate the criteria by which they can be taken to constitute the standard of perfection in industrial structure and performance.

An industry is traditionally deemed to be perfectly competitive if it possesses all three of the following attributes: (i) It is made up of a very large

number of firms, each of which provides so negligible a proportion of the industry's total output that no one firm's output decisions can have any discernible effect on price; (ii) the industry's products are perfectly homogeneous in the sense that no buyer distinguishes between the products of any two suppliers even in terms of any accompanying services, packaging or marketing procedures;[3] (iii) entry into and exit from the industry is totally unimpeded.

For the purpose of comparison with contestability analysis, the first of these three attributes is particularly noteworthy. It automatically rules out from the competitive category any industry in which, for example, scale economies make small firms relatively inefficient and consequently prevent their long-run survival in a free market. Because so many industries are characterized by technology that makes small enterprise completely impractical, the norm of perfect competition becomes not only unachievable, but, for many sectors of the economy, irrelevant. The concept of a perfectly contestable market is designed to provide a benchmark that applies in markets for which the concept of perfect competition is not very useful. An oligopolistic or even a monpolistic industry can be perfectly contestable if it is characterized by complete freedom of entry and exit—the last of the three attributes of perfect competition.

Formally, a market is defined to be perfectly contestable if no price in that market can be in equilibrium when its magnitude is such as to enable an entrant to undercut it and nevertheless earn a profit. Thus, a market that is protected by substantial entry barriers is clearly not contestable, because the barriers permit an equilibrium involving monopoly prices and monopoly profits. In the absence of barriers, those prices and profits would be undermined by entrants seeking to take advantage of the profit opportunity they provide. Thus, the matter can be looked at in a second and equivalent way. A market is perfectly contestable if firms can enter it and then, if they choose, exit without losing any of their investment. If this condition is satisfied, no prices set by the incumbents that offer profits to entrants can long endure. Thus, freedom of entry and exit are the key requirements of contestability.

The second version of the formal definition of a contestable market is tantamount to a requirement that there be no sunk costs.[4] A sunk cost, by

3. The product homogeneity attribute of perfect competition does not play any substantial role in the discussion that follows. It has been argued that heterogeneity of products can introduce a variety of inefficiencies. For example, it has been asserted that it tends to lead to a number of (slightly differing) products which exceeds anything justifiable in terms of costs and consumer benefits (as measured by the consumers' preferences) and that it leads firms to operate wastefully via unused capacity. The classic discussion of this subject is found in E. CHAMBERLIN, THE THEORY OF MONOPOLISTIC COMPETITION (1962).

4. Contestability theory draws heavily upon earlier work relating to barriers to entry originating

definition, is an outlay that cannot be recouped without substantial delay. If entry into a market requires a new enterprise to sink considerable amounts of capital, there must be violation of the requirement of perfect contestability that absolutely costless exit be possible. What is crucial is not the amount of capital that is required for entry, but the amount of this capital that is sunk.

Entry involving highly mobile capital, even if it is very substantial in quantity, may be followed by easy and rapid departure. Even if exit from the industry as a whole is difficult, mobility of capital may permit easy and rapid entry into and exit from particular markets in that industry. Similarly, industries using capital for which a strong second-hand market exists, or using capital that can readily be leased, are likely to exhibit the easy entry and exit characteristics needed for contestability. In light of these considerations, we can define the degree to which an industry is contestable. For we see that the smaller the share of investment that is composed of capital that is sunk, the more contestable that industry will be.[5]

A contestable market works most effectively if, in response to a profit-making opportunity, new firms can enter quickly, earn profits at least temporarily (before incumbents can institute countermeasures) and then leave without any loss of investment in sunk capital. This suggests that where incumbents can counterattack quickly, contestability will prevail only if hit-and-run entry can be carried out even more rapidly. It may appear that in the race between entry and retaliatory measures by incumbents the latter will generally prevail and so will preclude contestability; incumbents may be able to cut price almost instantly—as soon as entry

with J. BAIN, BARRIERS TO NEW COMPETITION (1956); it also draws from the literature on limiting pricing growing from the work of P. SYLOS-LABINI, OLIGOPOLY AND TECHNICAL PROGRESS (1962). Perfectly contestable markets contrast with the circumstances considered in the more usual oligopoly models in which both incumbent firms and new entrants must sink some costs. For markets in which sunk capacity is important and leads to strategic behavior when entry threatens or after it occurs, there is a well-developed literature, ably summarized in Dixit, *Recent Developments in Oligopoly Theory*, 72 AM. ECON. REV. 12 (1982).

5. Preliminary empirical evidence confirms that market behavior does indeed follow such a pattern—that is, the smaller the share of sunk outlays, the more closely the behavior of the firms in an industry follows the pattern to be expected in a contestable market. I. Kesides, Toward a Testable Model of Entry: A Study of the U.S. Manufacturing Industries (1982) (unpublished manuscript at Princeton University). Preliminary experimental evidence explores the conjecture that sunk costs weaken the discipline of contested markets, and finds that the disciplining power of market contestability is impressive. *See* Coursey, Isaac, Luke & Smith, *Market Contestability in the Presence of Sunk (Entry) Costs*, 15 BELL. J. ECON. (forthcoming, 1984). Preliminary experimental evidence also indicates that prices appear to be near competitive, rather than monopoly, levels when as few as two identical decreasing cost firms compete for a market large enough for only one of them. *See* Coursey, Isaac & Smith, *Natural Monopoly and Contested Markets: Some Experimental Results*, 27 J. L. & ECON. (forthcoming, 1984). These discussions suggest that the applicability of the standard of contestability is considerably wider than some of the initial discussions imply. *See, e.g.*, Dixit, *supra* note 4.

occurs or, for that matter, as soon as it threatens. But if entry into a market entails no sunk costs, a potential entrant has no reason to fear retaliation by incumbents, since it can leave the market without loss if such retaliation materializes. In many markets, rapid retaliation may not be possible; regulation, long-term contracts, or other impediments can slow the response of incumbents to entry. Moreover, a new firm can forestall retaliation by entering into contracts, before it actually opens for business, with customers it lures from incumbents.

B. *Virtues of Competitive and Contestable Markets*

Analysts have been attracted by the concept of perfect competition, despite its lack of realism and its inflexibility, because it has implications that can readily be used as standards of optimality for industrial performance. These include the preclusion of excess profits, the elimination of inefficient firms, the absence of cross subsidy, and pricing consistent with the allocation of resources available to the economy that is most efficient in serving the preferences of consumers. For purposes of comparison with contestability, it is useful to review each of these four attributes and to indicate why each is a necessary characteristic of a perfectly competitive industry.

Excessive profits are defined by economists as any long-run profits exceeding the cost of capital as determined by the markets for debt and equity. Such excess profits are eliminated in the long run by freedom of entry in a perfectly competitive industry. If the current cost of capital is twelve percent and a particular competitive industry offers a return of eighteen percent, new firms will be attracted into that industry, expanding outputs and driving down prices to the point where all excess profits have been squeezed out. The reason that inefficient enterprises cannot persist in a perfectly competitive industry is similar: cost inefficiencies invite replacement of the incumbents by entrants who can provide the same outputs at lower cost.

Cross subsidy, a problem that has long been of concern to regulators,[6] may be defined, roughly, as the sale by a multiproduct firm of some of its outputs at prices that are indefensibly low relative to their costs, with the resulting revenue shortfalls being offset by the charging of indefensibly high prices for other company outputs. For obvious reasons, cross subsidy is considered unfair both in its disparate treatment of the firm's customers and in its effects upon competing sellers of the products that are under-

6. For example, this issue has been the subject of more than a decade of hearings on telecommunications pricing before the FCC, with AT&T and its large business customers pitted against Western Union, MCI and other entrants. *See, e.g., In re* MCI Telecomm. Corp., 70 F.C.C.2d 666 (1979).

priced (i.e., those products which receive the internal cross subsidy). In perfectly competitive markets, cross subsidy is ruled out by the inability of firms to earn any excess profits. If no product of a firm can contribute more to revenues than its cost of capital, then that enterprise can have no source of funds with which to provide cross subsidies to any of its other outputs.

Of the four beneficial consequences of perfect competition, the most difficult to explain is the relationship between its pricing and efficiency in the allocation of the economy's resources. As already indicated, under perfect competition each firm is so miniscule a part of its industry that its output decision has no effect on price. The market-determined price therefore represents the addition to revenue from the production of another unit. The most profitable output of a good is, then, that output at which the marginal cost, the addition to the firm's total cost caused by the production of an additional unit, equals the given price.[7] Thus, perfect competition drives firms to equate marginal costs and prices—and that, according to economic analysis, is precisely the price behavior required for efficiency in the use of resources to serve consumers.[8]

These are the primary virtues of perfect competition that account for its widespread use as a criterion for industry performance. To what extent does perfect contestability share these attributes?[9] To begin with, perfect contestability precludes both excess profits and inefficient firm operations in the long run, for much the same reasons that perfect competition precludes them. Should either phenomenon arise temporarily, it would make possible profitable operation by firms charging prices below those of the

7. Thus, if the market price of X is \$10, while its marginal cost (including the cost of the additional capital required) is \$9.25, the firm can add to its profit by expanding its output, for it gains \$.75 on each additional unit produced. Gradually, however, diminishing returns will increase marginal cost and erode the profitability of further expansion until a marginal cost equal to the \$10 price is attained, and it pays the firm to set its output there.

8. The reasoning underlying this standard result is fairly straightforward. If a consumer purchases a unit of good X, the cost his purchase causes, i.e., the value of the resources used up in meeting this demand, is by definition precisely the marginal cost of X. Thus, if the price of each good is equal to the marginal cost of that item, the prices consumers pay are the same as the costs caused by their purchases. Money cost then becomes a perfect proxy for the real social cost incurred in providing a unit of a good to the consumer. Then, if consumers use their *money* resources optimally from their own point of view, i.e., in a way which best serves their preferences, they will automatically be using the economy's *real* resources for the satisfaction of consumer desires as efficiently as is possible. In contrast, if the price of X were low relative to its marginal cost while the reverse were true of good Y, then consumers would be attracted to buy more of X and less of Y than the true economic costs call for. For example, suppose X and Y are perfect substitutes for consumers but the marginal cost of X and Y are \$5 and \$4, respectively. If, however, the price of X is \$5 while that of Y is \$5.50 consumers will spend their money efficiently by purchasing X in preference to Y. But that, clearly, is not compatible with efficiency in the use of the economy's resources.

9. For a full examination of the issue with all its technicalities, see W. BAUMOL, J. PANZAR & R. WILLIG, *supra* note 1, at 191-345. A more intuitive discussion is provided in Baumol, *Contestable Markets: An Uprising in the Theory of Industry Structure*, 72 AM. ECON. REV. 1, 4-5 (1982).

incumbents. Since entry into a perfectly contestable market is costless, this opportunity for short-term profits would attract entry—and that in turn would soon force down prices and profits. Furthermore, with no possibility of excess profits on any of a firm's products, there can be no source of cross subsidy. Thus the first three virtues of perfect competition are realized in perfectly contestable industries, and these traditional concerns of regulators become groundless.

The possibility of economies of scale in perfectly contestable industries makes the evaluation of the fourth criterion—the equation of price with marginal cost—more complex. The argument for the case of perfect competition is inapplicable in the presence of economies of scale: efficient sized firms are large enough to affect price by their output decisions. However, if scale economies, though significant, do not create a natural monopoly, perfect contestability ensures marginal cost pricing. Moreover, even in the case of natural monopoly, perfect contestability ensures that the lowest price consistent with the continued provision of the good is charged.

Consider the case in which scale economies do not result in natural monopoly. Then maximum efficiency is consistent with two or more firms operating in the industry. A price exceeding marginal cost will create an opportunity for profitable entry. The entrant can (slightly) undercut the incumbents' price, and sell slightly more units than some of the incumbents. The profit on the additional sales must more than cover the reduction in profits resulting from the price reduction on the preceding units. Similarly, if price is below marginal cost, an entrant can offer a slightly lower price than the incumbents, and sell slightly fewer units. The additional profits from the reduction in output must more than cover the decrease in profits resulting from the decrease in price. Thus prices above or below marginal costs invite entry, and therefore price will be in equilibrium only if it equals marginal cost.[10]

In natural monopoly, a single producer can achieve a lower total cost than can any group of firms and total costs are usually large relative to marginal costs.[11] Only one efficient sized firm can remain. Moreover, marginal cost pricing is likely to involve costs that exceed revenues, and so

10. A problem can arise if products in a perfectly contestable industry are heterogeneous, each supplier offering his own special brands with their own special features. However, it can be shown (through an argument similar to that made above) that if each variant is sold by at least two different suppliers, perfect contestability will lead to marginal cost pricing. *See* W. BAUMOL, J. PANZAR & R. WILLIG, *supra* note 1, at 314-21, 329-45.

11. It is tempting to compare marginal costs with *average* costs but in a multiproduct firm (and virtually all firms in reality are multiproduct enterprises) average costs cannot even be defined, because of the outlays which are almost always incurred in common on behalf of several of the firm's products. To give the trivial but standard example, there is no way to determine what part of the salary of the president of the firm is "caused" by the supply of product *A* rather than product *B*, and therefore constitutes a legitimate portion of *A*'s average cost.

no such firm can afford such a pricing policy. However, contestability does ensure that a natural monopolist will be able to prevent entry only if it sets the lowest prices consistent with the financial viability of the firm. Otherwise an entrant could slightly undercut his prices, taking the entire market for himself, and earn a normal profit. Similarly, costless entry precludes inefficient operation. Moreover, with no above normal profits on any goods, there is no opportunity for cross subsidy, even by a contestable monopoly.

This discussion reveals the reason contestability analysis departs from the tradition which classified scale economies as a barrier to entry. Contestability analysis defines an entry barrier as something which provides incumbent firms sufficient protection from entry so that they can obtain above normal profits or exhibit other forms of unacceptable performance. But we have seen that perfect contestability guarantees the absence of excess profits, inefficiencies and cross subsidies even in the presence of scale economies. It can also be shown that even in the presence of a natural monopoly, contestability rewards the firm for selecting the prices that are most efficient given the requirement of solvency of the enterprise. Thus, scale economies are not a source of undesirable performance in a contestable market and cannot be considered a form of entry barrier.

But scale economies do affect the usefulness of contestability relative to perfect competition as guides for policy. We have seen that to be perfectly competitive an industry must be populated by a large number of miniscule firms. But even if each of these dwarf enterprises operates with exemplary efficiency, the overall result may be inefficient. Suppose, for example, that the industry's technology provides substantial economies of scale or what has come to be called *economies of scope*—i.e., economies that derive from the simultaneous production of several goods or services.[12] In these circumstances, large and diversified firms may be able to supply goods far more cheaply, in terms of resources used, than can the many small firms of perfect competition.

12. Economies of scope are defined formally as follows: Let $C(x,y)$ be the cost of production by a single firm of quantity x of some good, X, and quantity y of another commodity, Y. Let $C(x)$ be the cost of producing x if it is done by a completely specialized firm and $C(y)$ have an analogous connotation. Then production of X and Y is characterized by economies of scope if $C(x,y)$ is less than the sum of $C(x)$ and $C(y)$. That is, economies of scope are present if the multiproduct firm's production of x and y together incurs a cost lower than the sum of the cost of x alone, when produced by a specialized firm, and the cost of y when produced by another specialized enterprise. For example, a telecommunications network can serve many routes more efficiently than it can only a few. If calls from New York to Los Angeles have just been transmitted via Chicago and the Chicago portion of the route encounters a surge of traffic which strains its capacity, the New York-Los Angeles calls can be (and are) rerouted, say, via New Orleans. Thus, simultaneous provision of service to these four regions makes it possible to operate with lower capacity in both New Orleans and Chicago. For similar reasons shoe factories usually produce footwear of different sizes and styles, automotive firms usually produce both cars and trucks, etc.

This type of inefficiency is referred to as *inefficiency in industry structure*, as distinguished from inefficiency in the operations of individual firms. We see now that perfect competition, with its requirement that all firms be small, may be inconsistent with efficiency in industry structure; when that is the case and markets are free, perfect competition will be unable to survive. In an industry characterized by perfect contestability, however, industry structure *must* be efficient. If an industry is perfectly contestable and its outputs can be produced more cheaply by four firms than by three or six or any other particular number of enterprises, then in the long run that industry must indeed be composed of four firms. A two-firm or a nine-firm industry structure cannot survive the market pressures introduced by perfect contestability.

To see why this is so, suppose that by happenstance the hypothetical industry contains nine enterprises even though four-firm production is least costly. Some of those nine firms can be depended upon to seize the cost-reduction opportunity offered by the availability of economies of scale and scope. These economies will permit the more enterprising firms to undercut the prices of rivals who are slower to grasp the expansion opportunity. Costlessness of exit will then make it easy for those laggard rivals to leave the industry. If no incumbent in the industry is sufficiently enterprising to take advantage of the opportunity to undercut prevailing prices, outsiders sufficiently large to benefit from the economies of scale and scope will be happy to do so, since in a contestable industry they bear neither entry cost nor risk. In due course, surplus firms will be driven from the industry, and a four-firm structure that minimizes the total cost of its bundle of outputs will emerge. Such a four-firm industry may be perfectly contestable, but it can hardly qualify as perfectly competitive.

This, then, is the sense in which it can be said that while perfectly competitive and perfectly contestable markets are both ideals, the latter is more ideal than the former. After all, one must be tempered in one's praise of the many-firm structure of perfect competition in those cases in which the availability of economies of scale and scope means that an oligopoly structure can (perhaps) achieve far lower costs and offer far lower prices to consumers.

The perfectly competitive structure is simply not attainable in a broad group of industries. The idea of transforming the automobile, steel, or telecommunications business into an industry composed of a huge number of tiny enterprises is absurd on its face. That is why perfect contestability is a standard of structure and performance that is more pertinent than pure competition given the character of modern technology.

C. *Some Features of Contestable Markets*

Several attributes of a contestable industry are particularly significant for regulatory policy and therefore merit explicit discussion. We will begin by discussing several of the requirements that must be satisfied in order for a market to be contestable; then we will consider some of the implications of contestability for market behavior.

Freedom of exit is a crucial ingredient of contestability. In a contestable market, freedom of exit is merely the obverse of freedom of entry. Any impediment to exit by definition increases the riskiness, and hence the real cost, of opening for business. A potential entrant will hesitate long before embarking on an enterprise from which it will be difficult to withdraw if his entry proves to have been a mistake. This means that the traditional resistance of regulators to exit—for example, their refusal to permit the abandonment of unprofitable routes by railroads or airlines—is hardly without cost to the economy. However laudable the motivation for opposition to exit—be it the preservation of service to isolated consumers, the safeguarding of jobs, or the maintenance of tax bases in financially troubled communities—the fact is that it has an unintended adverse consequence: preclusion of, or restraints on, exit discourage entry and thereby reduce the competitive threat posed by the availability of potential entrants. The moral is that when subsidies do serve the public interest, regulators should consider ways to provide those subsidies directly rather than attempting to encourage cross subsidy by making exit difficult.

This immediately suggests a second necessary feature of contestable markets: the availability of a pool of potential entrants able to respond quickly to an entry opportunity and to choose the timing, place, and manner of entry that best suits the circumstances. It is their *threat* that disciplines incumbents and forces them to serve consumers efficiently.[13] A regulatory process in which lengthy hearings and evidence of public convenience and necessity are prerequisites to entry is precisely what is not required.[14] Contestability is also subverted by the regulatory custom that requires potential entrants to commit themselves well in advance to the timing and manner of their proposed entry.[15] Contestability requires that firms have what can be described as *standby* authority to enter a

13. The role of potential entry and its threat was first emphasized by Bain, and he, too, maintained that it has generally not been given adequate attention. J. BAIN, *supra* note 4.

14. For example, hearing processes at the CAB tended to take from one to four years to complete, and even then almost no new authorizations were granted for routes that already had two or more authorized carriers. Normally, in a route case at most one new carrier was selected for the route. *See* Bailey, *Deregulation and Regulatory Reform of U.S. Air-Transportation Policy*, in REGULATED INDUSTRIES & PUBLIC ENTERPRISE: EUROPEAN & UNITED STATES PERSPECTIVES 29, 29 (1980).

15. For example, CAB rules also required proposed operating schedules to be submitted at the beginning of a route case.

market—authority that need not be exercised so long as good performance by incumbents precludes profit opportunities for entrants, but that can be used quickly when unsatisfactory incumbent performance offers entrants the prospect of profit.

In addition to these two crucial elements of contestability, there is a third feature that facilitates, but is not absolutely necessary to, contestability: sluggishness in the responses of incumbents, particularly the pricing responses, to entry. This does not mean that incumbents should be prevented from competing fully and effectively by adjusting prices. The entire purpose of the competitive process from the viewpoint of the general welfare is that it forces firms to offer low prices and to provide service and products of high quality. Contestability theory does suggest, however, that regulation-induced lags in pricing may well be salutory. While slowness in incumbents' pricing responses to entry is conducive to contestability, it is by no means essential. Potential entrants can and do sometimes make binding contracts with their future customers, and if such contracts can be agreed upon quickly, the fact that entry takes a longer time becomes irrelevant. Once the contract is signed, a retaliatory price reduction by incumbents will have lost its sting.

We turn now to some pertinent consequences of contestability for regulation. One key characteristic of contestable industries is the dependence of their structure—including not only number of firms but also the degree of their integration, the number of different items in their product lines, and the dispersion in the sizes of enterprises—upon the forces of the market. This efficiency attribute of long-run equilibrium in a perfectly contestable market means that any structure compatible with equilibrium must offer the lowest cost that is attainable. This implies that regulatory attempts to influence the structure of an industry, perhaps seeking to increase the number of firms it contains, are often doomed to failure. Newly introduced enterprises either will not survive or will replace some incumbents, but in the long run the number of firms will be unaffected. In such circumstances, regulators are all too often tempted to keep firms alive by subverting competitive pressures—specifically, by establishing what amounts to a cartel, in which each enterprise is protected from the competition of the others. But an arrangement of that sort is a monstrosity that keeps up the appearance of competition by assuring the survival of firms as an end in itself, by completely undermining the competitive process and imposing a heavy cost upon the consuming public.

A second implication of contestability for regulation relates to the good behavior that perfect contestability imposes upon business firms. As we have seen, the discipline imposed by the possibility of hit-and-run entry precludes pricing above marginal costs, excess profit, inefficient operation

or structure, and cross subsidy, even in an oligopoly. If a perfectly contestable industry is a natural monopoly, the last three conditions are met and the firm must charge as low a set of prices as long-run financial viability permits. Thus, when an industry is contestable, not only is regulatory intervention designed to influence industry structure ineffective, but it is also unnecessary for the protection of the interests of consumers.

Finally, we observe from what has just been said that in a contestable industry the absence of entry should not be taken as a sign of predatory behavior by incumbents, but, on the contrary, as an indication of good behavior on their part. In a perfectly contestable industry, as we have seen, entry can be prevented only by performance that satisfies competitive standards. Thus, if an industry is reasonably contestable a regulator should certainly hesitate before deciding that a history without entry constitutes grounds for intervention.

One other issue should be noted at this point: Since *perfect* contestability is highly improbable in reality, it is necessary to consider the state of affairs that is likely when the requirements of contestability are fulfilled only approximately. Unfortunately, since the entire analysis is so new, the case of imperfect contestability and the attributes of a "workably contestable" market are only now being explored.

In a recent note, Marius Schwartz and Robert J. Reynolds[16] of the Department of Justice argued that slowness of entry can cause serious problems for performance even if the lag is only moderate. They assert that, in such circumstances, rather than acting like perfect competitors and earning zero economic profits, incumbents will find it rewarding to adopt prices that yield them monopoly profits during the period before entry occurs—and then exit gracefully when the entrant opens for business, undercuts the incumbent, and takes the market over. These authors offer several other scenarios designed to show that slight departures from the requirements of contestability can cause large deviations from contestable behavior.

In reply, Baumol, Panzar and Willig[17] undertake to show that Schwartz and Reynolds do not use the right criterion of proximity to contestability. In their view, it is the magnitude of sunk costs, rather than the entry lag, that is the crucial issue. It is clear that as sunk costs approach zero the risks associated with entry also become negligible because an entrant who finds in retrospect that his entry decision was an error can then

16. Schwartz & Reynolds, *Contestable Markets: An Uprising in the Theory of Industry Structure: Comment*, 73 AM. ECON. REV. 488 (1983).

17. Baumol, Panzar & Willig, *Contestable Markets: An Uprising in the Theory of Industry Structure: Reply*, 73 AM. ECON. REV. 491 (1983).

pick up *almost* all of his marbles and depart with commensurately little loss. Thus, if sunk costs are small but not zero, the discipline exercised by the threat of potential entry remains potent.

The meager empirical evidence also supports this view. But it must be granted that results on this subject are still highly preliminary and do not yet lend themselves to confident generalization. Much of what can be concluded about this issue now derives from particularized observations, such as those in the discussion below of recent regulatory issues. It may be added, however, that we are as unsure about approximations to perfect competition as we are about approximations to perfect contestability. The literature on workable competition is suggestive and illuminating, but it rests neither on rigorous formal analysis nor on a clear-cut body of empirical evidence.

II. The Implications of Contestability Theory for Regulatory Policy

We have argued that perfect competition is neither an attainable nor a desirable benchmark for industries in which economies of scale or scope are substantial; in such cases, attempts to approximate perfect competition may in fact be highly inefficient. However, the divergence of such industries from the patterns of perfect competition does not justify the sorts of regulation that have traditionally been imposed. Regulators should seek policies that promote contestability. If an industry behaves as if it is contestable, most of the benefits of perfect competition can be obtained without government intervention. In short, our position is that the equilibrium of a contestable market is often a better standard for public policy than the competitive model, particularly in the presence of economies of scale and scope.

Short of doing everything possible to foster contestability, regulators should certainly cease doing those things that work against it. Direct regulatory attempts to impede entry or exit or to interfere with the timing or manner of entry must, at the very least, be questioned severely. Moreover, regulators should keep their eyes open for entry barriers erected by firms and should take steps to discourage the maintenance of those barriers.[18]

In addition to entry barriers introduced artificially by regulators or incumbent firms, there are in many industries what may be described as "natural" barriers, i.e., barriers that arise out of technological circumstances. For example, the technology of an industry may require heavy sunk investments on the part of entrants, as we have seen. An investment

18. We will not pursue this subject here because it has been raised and discussed long before the advent of contestability analysis, and this analysis has, so far, not shed any new light on that part of our subject.

that cannot easily be moved elsewhere is an impediment to exit, which, as has been shown, is in turn a prime obstacle to entry. However, as one critic has noted to us in conversation, "One cannot regulate away the need to sink costs." What, then, can be done to weaken the barrier to entry that is found when entry requires heavy sunk investments?

In these cases, regulators are just beginning to experiment with new methods to ensure that no excessive profits are earned from sunk-cost facilities. Rather than relying exclusively on traditional rate and entry regulation, they have turned to two rather novel approaches. The first of these entails government intervention to ensure equal access to the sunk facility. If the facility is privately owned, the government requires that all firms seeking to use the facility be given access to it, that the access price be reasonable,[19] and that all users be charged the same price. If the sunk facility is in the hands of a local public authority, then that authority is encouraged not to discriminate among private users in its access policies.

The second approach is to isolate the sunk investments, leaving a relatively contestable part of the industry's operations to be controlled by market forces, while the portion with substantial amounts of sunk capital is regulated or even operated by the public sector. Thus, some new legislation and some regulatory decisions are characterized by a flexible case-by-case approach, in which markets subject to strong competitive pressures from substitute services and markets in which technology does not require heavy sunk costs are freed from traditional regulatory constraints and are permitted more open entry and more flexible pricing.[20] Those segments of an industry that have large sunk costs or for which there is a problem of "nonsustainability," that is, absence of an equilibrium,[21] must continue to

19. See the discussion of airport access, *infra*, text accompanying notes 44-45, or the discussion of access to local telephone systems, *infra* at Sec. II.D. The lesson that continued regulation to maintain open access may be appropriate holds even more strongly for industries that are characterized by large sunk costs and long lead times, such as the electric power industry. For example, it is important to design and apply criteria for guaranteeing access to transmission systems and power pooling activities. *See* P. JOSKOW & R. SCHMALENSEE, MARKETS FOR POWER: AN ANALYSIS OF ELECTRIC UTILITY DEREGULATION (1983).

20. See the discussion of the Fresh Fruit & Vegetable Decision, *infra*, text accompanying note 28, or the discussion of the AT&T settlement, *infra*, text accompanying note 66. Another example is the 1979 Amendments to the National Health Planning and Resources Development Act of 1974, Pub. L. No. 96-79, 93 Stat. 589 (codified in scattered sections of 42 U.S.C.), which attempts to distinguish between, and treat rather differently, in-patient institutional services and noninstitutional and out-patient services. For in-patient services, continued regulation is called for and the utility of contestability analysis is not great because of both sunk cost problems and third-party reimbursement mechanisms. However, contestability analysis is consistent with freeing from entry regulation those portions of the noninstitutional and out-patient services that are characterized by fixed, but no sunk, costs. *See* J. GELMAN, COMPETITION AND HEALTH PLANNING (F.T.C., 1982).

21. In the single product case, no equilibrium is possible if a natural monopoly firm is producing in a rising portion of its average cost curve. The disequilibrium occurs because any group of customers who together demand an amount of the product equal to the level at the minimum point on the average cost curve can supply themselves at a lower price. Thus, they have an incentive to leave the

be regulated.

We now turn to a description of reform measures undertaken in the transportation and communications industries during the late 1970's and early 1980's. We focus in each case upon two or three aspects of reform that reflect the features of contestability that have been emphasized. Our descriptions are brief rather than thorough. They outline a framework for policy which is consistent with contestability theory. The major lesson that emerges from this analysis is that contestability theory provides no mechanistic prescriptions or inviolable rules for regulators or for authors of regulatory reform legislation. However, it does offer substantial insights that can strengthen the effectiveness of their work.

A. *Reform of Railroad Regulation*

The sunk cost and longevity of railroad capital may suggest that the railroad industry is one in which contestability analysis cannot conceivably apply. However, the railroad industry is more contestable than has been traditionally acknowledged, because there is strong competitive pressure from other modes of transportation—such as trucking—on the rates charged for shipment of a wide variety of commodities. Contestability analysis tells us that even in markets in which sunk costs are substantial, pricing power may be held in check by the availability of substitute suppliers whose cost structure is compatible with contestability. In these circumstances, the theory suggests that rate regulation is not required.

The railroad legislation passed in 1976, the Railroad Revitalization and Reform Act (4R Act),[22] is in harmony with this suggestion. From the standpoint of contestability, the most pertinent provision of the legislation is one that offers a railroad freedom in pricing where there is no evidence that it holds a position of "market dominance."[23] The 4R Act provided that market dominance "refers to an absence of effective competition from other carriers or modes of transportation, for through traffic or movement."[24] The Interstate Commerce Commission (ICC) has listed four

natural monopoly. The absence of a stable price can occur in the multi-product case as well, and is particularly likely to occur when there is strong demand substitutability and product-specific scale economies. In public policy terms, lack of a stable price equilibrium in the face of open entry and exit may require entry regulation. *See* Faulhaber, *Cross-Subsidization: Pricing in Public Enterprise*, 65 AM. ECON. REV. 966 (1973); *see also* BAUMOL, PANZAR & WILLIG, *supra* note 1.

22. Pub. L. No. 94-210, 90 Stat. 31 (codified as amended in scattered sections of 45, 49, 15 & 31 U.S.C.).

23. The 4R Act prohibited the ICC from deciding that a rate was excessively high without a finding that the carrier filing the rate had market dominance. *Id.* § 202(b), 49 U.S.C. § 10,709. In 1980, Congress limited the ICC's jurisdiction to determine that rates are reasonable to those rates established by rail carriers which the ICC finds have market dominance. Staggers Rail Act of 1980, Pub. L. No. 96-448 § 201(a), 49 U.S.C. § 10,701a (Supp. V 1981).

24. 49 U.S.C. § 10,709(a).

types of competitive checks that are to be considered in determining whether there is market dominance over the transport of a particular product: intramodal competition, intermodal competition, geographic competition (the ability to transport the product to or from a different location) and product competition (the availability of substitutes for the product).[25] In principle such an approach is entirely consistent with the implications of contestability analysis. When an industry can be segmented into independent components, it is desirable to free from regulation those parts of the industry in which competing firms lack market power.

The 4R Act also granted the ICC the authority to exempt the rail carriage of certain goods and passengers if the rail carriage was of "limited scope" and regulation was not otherwise desirable.[26] The Staggers Rail Act of 1980 extended this authority, allowing the ICC to exempt a movement if it finds that rate regulation is not needed to carry out the transportation policy of the Act, and that either the carriage is of "limited scope" or that regulation "is not needed to protect shippers from the abuse of market power."[27] Perhaps the best-known recent example of such an exemption is the ICC's decision to permit railroads total freedom of pricing in the transportation of fresh fruits and vegetables.[28] This action also accords well with the prescriptions of contestability theory as was recognized by economist Darius Gaskins, who was Chairman of the ICC when the exemption was passed.

It is noteworthy that the 4R Act did not automatically extend freedom from regulation to segments of the industry in which a railroad was alleged to dominate the market. Consider, for example, coal transport. Leaps in the price of petroleum, first in the mid-1970's and then again in 1979 and 1980, caused the demand for coal to rise substantially. Railroads are the chief transporters of coal to major coal-using facilities, such as electric power plants. Once these plants have been constructed, their relocation is largely precluded. It is clear that Congress was right in continuing regulation of rates when the ICC finds that a carrier has market dominance.[29]

25. *See* Market Dominance Determinations & Consid. of Product Competition, 365 I.C.C. 1, 118 (1981).

26. 49 U.S.C. § 10,505.

27. *Id.* § 10,505(a).

28. Rail Gen. Exemption Auth., 361 I.C.C. 211 (1979).

29. *See* 49 U.S.C. § 10,701a(b)(1). For a more thorough discussion of ICC activity during this period, see Gaskins & Voytko, *Managing the Transition to Deregulation*, 44 LAW & CONTEMP. PROBS. 9 (1981); *see also* A. FRIEDLAENDER & R. SPADY, FREIGHT TRANSPORT REGULATION (1981); Eads, *The Reform of Economic Regulation*, in AMERICAN ENTERPRISE INSTITUTE, TELECOMMUNICATION AND TRANSPORTATION (1982).

The theory of contestability suggests that in such cases policymakers should consider ways to encourage potential competition from other sources. But the case of the railroads illustrates that this must be done with care and that the most obvious paths may be beset by hidden perils. For example, if railroad *A* is required to permit competing railroad *B* to lease space on any of *A*'s tracks—that is, to grant *B* what are known as trackage rights—the arrangement would appear to enhance competition by permitting *B* to bid for shipments along routes that parallel *A*'s but can reach their ultimate destination only by traversing some of *A*'s track. This is indeed so if the price for trackage rights is settled voluntarily in a free competitive market. However, if regulators force the provision of trackage rights at a noncompensatory price, this action will amount to a subsidy to railroad *B* that effectively drives *A* from the field and undermines competition and efficiency rather than enhancing them.

Similarly, some proponents of enhanced competition have proposed that the right of eminent domain be granted to coal slurry pipelines in order to force railroads to permit the pipelines to cross rail property.[30] It is asserted that this will increase greatly the competitiveness of coal transportation. But here, too, there is a pricing complication. Environmental groups claim that the pipelines will make heavy use of scarce water supplies and that disposal of waste products from the pipelines will exact a heavy pollution cost. The pipelines should be forced to bear the full costs of water use and waste disposal through direct charges or other measures. Otherwise, heavily subsidized entities will be pitted against firms that are required to cover their own costs, resulting in inefficiency and environmental degradation rather than real competition.

Whatever the true facts in these matters, the moral is clear. The enforced introduction of competition must not be accompanied by artificial prices for incumbents or entrants; otherwise, more harm than benefit may flow from it, and no contribution to contestability will result.

B. *Aviation Deregulation*

The Airline Deregulation Act of 1978[31] rested upon such a degree of confidence in the inherent structural competitiveness of the domestic U.S. airline industry that it went further in deregulating than any other piece of recent legislation. Regulatory barriers to entry of the type that had

30. *See, e.g.*, Tarlock, *Western Water Law and Coal Development*, 51 U. COLO. L. REV. 511, 538 (1980). One pipeline company has won right of way disputes with railroads in the Eighth and Tenth Circuits. Energy Transp. Sys., Inc. v. Union Pac. R.R. 619 F.2d 696 (8th Cir. 1980); Energy Transp. Sys., Inc. v. Union Pac. R.R., 606 F.2d 934 (10th Cir. 1979).

31. Pub. L. No. 95-504, 92 Stat. 1705 (codified as amended in scattered sections of 49 U.S.C.).

been favored by the Civil Aeronautics Board (CAB) were removed within three years of its passage.[32] Under the Act, the CAB can no longer block new jet carriers from entering the industry, nor can it conduct lengthy route cases to decide which additional carrier will be permitted to serve a particular pair of cities.[33] The new legislation recognizes the benefits of permitting potential competitors to respond to profit opportunities by entering markets freely. Even though the number of actual competitors in most markets might not change very much as a consequence, the Act gave airline managements complete freedom in the structuring of their route networks,[34] relying on this freedom as an adequate check on market power. The Act also provided for complete freedom of pricing, effective one year after entry became unimpeded.[35] The only exception to the deregulatory tone was for air services involving small communities: to avoid curtailment of these services, the Act provided direct subsidies for a ten-year period.[36]

In some ways the airline industry presents a particularly close approximation to contestability. As discussed above, the more mobile the capital and the smaller the sunk costs involved in an industry, the more that industry approaches perfect contestability. The major component of capital equipment in the airline industry, the airplanes themselves, can readily be moved from market to market. Such items of equipment are, in Alfred Kahn's words, "marginal costs with wings." This is true even though, because of technological economies of scale with respect to aircraft size, the majority of U.S. city-pair markets are natural monopolies (and so are likely to be served by only one carrier even under free entry) and all markets are likely to show high concentration.[37]

The evidence on route networks since deregulation corresponds well to the theory. Because of the economies of aircraft size, the cost of accommodating a passenger in an otherwise empty seat is quite small. Airlines thus have a strong incentive to establish hub-and-spoke operations. Flights from various origins arrive at an intermediate point where passengers change planes to proceed to their ultimate destinations. By combining passengers with different origins and destinations, a carrier can increase the

32. 49 U.S.C. § 1551(a).

33. *Id.* § 1371(d)(7)(A).

34. *Id.* § 1302(a)(4).

35. *Id.* § 1551(a)(2)(B).

36. *Id.* § 1389. For a description of the events leading up to the Airline Deregulation Act and a more complete analysis of its features, see S. BREYER, REGULATION AND ITS REFORM 197-221 (1981).

37. As described in D. GRAHAM & D. KAPLAN, COMPETITION AND THE AIRLINES: AN EVALUATION OF DEREGULATION 54 (1982) (CAB Staff Report), the average number of large aircraft operators providing nonstop service was less than four carriers even in the densest markets (those having more than 500 passengers per day).

average number of passengers per flight and thereby reduce costs. Essentially, the broader scope of operations lets the carrier take advantage of the economies of scale in aircraft size. Because of these advantages of hub-and-spoke operation, both the trunk and local service carriers[38] have increased their use of hubbing in the deregulation period when entry has been free. For example, in 1978, only three of the sixteen regulated airlines had twenty percent or more of their total domestic departures out of their leading city; by 1981, the number had increased to ten airlines. On the other hand, open entry has not led to extreme proliferation of carriers on routes. The average number of carriers per route has increased, but the increase has been moderate. For instance, in the post-deregulation period, approximately two carriers per route serve in the moderate-density, short-haul markets, whereas three to four carriers per route serve in the densest markets.

The condition of contestability theory that incumbents' prices must be relatively "sticky" is not met in aviation. In many cases, incumbent airlines respond immediately to meet a new competitor's lower fares. So it may not be surprising that the evidence on pricing policy since deregulation is more complex. In theory, after a transition period, potential rather than actual competition should serve to police markets. In particular, if all airlines face the same market demands and have access to the same productive techniques as those available to incumbent firms, actual entry should not be needed to limit prices, and price wars and related strategic behavior by incumbents should not be observed. We know that during the first year or two after deregulation, trunk carriers were held to lower price ceilings than local service carriers. Bailey and Panzar[39] showed that despite the presence of economies of density in city-pair airline markets, potential competition by trunk carriers was effectively policing the pricing behavior of local service carriers in their long and medium-haul routes. However, in trunk markets during that period, and in virtually all markets since then, actual (as distinguished from potential) competition has been found to play a significant role.

Graham, Kaplan and Sibley have concluded that the presence of newly

38. Sixteen trunk carriers were certificated in 1938; this number had shrunk through mergers to ten carriers by the time of passage of the Airline Deregulation Act. Trunk carriers are large jet aircraft operators serving dense city-pair markets. In contrast, local service carriers are specialty carriers, certificated by the Board in the mid-1950's to provide subsidized feeder service to small communities in nonoverlapping regions. By 1978, these carriers, too, were largely jet operators, but typically used smaller, two-engine jets rather than the three-engine and four-engine jets typical of the trunk carriers. For a history, see E. BAILEY, D. GRAHAM, & D. KAPLAN, DEREGULATING THE AIRLINES (forthcoming, 1984).

39. Bailey & Panzar, *The Contestability of Airline Markets During the Transition to Deregulation*, 44 LAW & CONTEMP. PROBS. 125, 134-44 (1981).

certificated carriers has a substantial effect on fares in the markets they serve.[40] They also found that fares are higher in markets served by the four airports[41] in which there is a shortage of landing slots and higher also in markets in which concentration is quite pronounced. Price wars have broken out on the denser routes normally flown by three-engine and four-engine jets. Moreover, incumbent carriers appear to be calculating whether or not they would be better off if they fail to match the lower fares of new entrants and learn to coexist with them rather than matching these fares in order to try to drive out the less-established enterprises. This is, of course, a behavior pattern one would expect from rival oligopolists in the standard analysis, not from players in a perfectly contestable world.

This evidence suggests that the contestability benchmark does not fully hold sway in the first years after deregulation. Why is the industry characterized by fierce rivalry rather than by quiet long-run equilibrium? The pure theory of contestable markets is an analysis of equilibrium conditions, just as the pure theory of perfect competition is. In the current reality in aviation, many of the assumptions underlying stationary equilibrium theory simply are not holding true. For example, the route network was closed to free entry for forty years. It is not reasonable to expect an instant adjustment to a deregulated equilibrium. Instead, market shares of major groups of carriers have been shifting rapidly, with the large trunk carriers losing market share to local service carriers and new entrants.[42] The doubling of fuel prices between 1978 and 1981 has had a profound effect on optimal aircraft deployment, and has meant that two-engine jets have become relatively more efficient than three-engine and four-engine jets. There is thus substantial excess capacity in three-engine and four-engine jets, and undercapacity in two-engine jets. Economic theory suggests that undercapacity will lead to higher prices until more planes can be brought on line. The expected response to overcapacity is price wars. When the excess capacity is substantial, prices will be near variable costs rather than full marginal costs (including the cost of capital). This signals to the investment community that additional capital is not required in the area.

Another standard assumption of equilibrium theory is that all players in the market have the same cost structures. This is not currently true in aviation. Costly labor contracts and associated restrictive work rules nego-

40. D. GRAHAM, D. KAPLAN & D. SIBLEY, EFFICIENCY AND COMPETITION IN THE AIRLINE INDUSTRY (1982) (CAB Staff Report).

41. The four are Washington's National Airport, Chicago's O'Hare Airport, and New York's Kennedy and LaGuardia Airports.

42. *See* E. BAILEY, D. GRAHAM & D. KAPLAN, *supra* note 38.

tiated during forty years of regulation bind the older trunk carriers; most of the new entrants, however, are not unionized. Thus, new entrants have a cost advantage and can earn profits at prices that are not compensatory to incumbent carriers. A third, and apparently very important, influence impeding contestability is the PATCO job action, which restricted entry into—and has provided an incentive to avoid exit from—major airports. Slots have generally been treated as a vested and non-tradable right of incumbent carriers, with entry by competing carriers largely precluded because of the freeze in capacity.[43]

The factors just discussed have impeded contestability. Given the inherent ability of airlines to move in and out of markets relatively unconstrained by sunk costs, these markets should exhibit more of the character of contestability before a great deal of time has passed. But it is unlikely that perfect contestability will be achieved over the short run since changes in labor contracts and fleet configuration cannot be carried out quickly.

Before leaving this analysis of the aviation industry, two additional implications of contestability analysis should be noted. First, even in aviation, sunk costs are present—particularly at airports. Theoretically, if a particular airline were permitted to own an airport, then it could obtain the monopoly rents associated with that airport, through the prices charged to its passengers or to the other carriers permitted to use the facilities at that airport. Thus, the prevention of control and ownership of airports by particular airlines is important for the contestability of markets in the industry. In general, rules of access to airports should be given careful consideration by policymakers. Accordingly, it is a matter for concern when local airport authorities attempt to deal with slot or noise constraints by banning new entry while permitting incumbent carriers to expand at will.[44] Another matter for concern is the appropriateness of long-term lease arrangements that allocate airport space to particular carriers and that give these carriers the power to determine when, to whom, and at what price to sublease space to their competitors.[45] A similar concern over newcomers' access has arisen with respect to computer reservations. In some regions of the country, travel agents predominantly use a single

43. *See* Grether, Isaac & Plott, *The Allocation of Landing Rights by Unanimity Among Competitors*, 71 AM. ECON. REV. PAPERS & PROC. 166, 167 (1981).

44. For example, San Diego sought in 1979 to exclude new airlines but did not plan to restrict new access by existing incumbents. After the FAA and CAB intervened the airport authorities withdrew their proposal. Orange County's John Wayne Airport had a similar plan which is currently under challenge in federal court. Pacific Southwest Airlines v. County of Orange, No. 81-3248 (C.D. Cal. 1983).

45. *See* AIRPORT ACCESS TASK FORCE, REPORT AND RECOMMENDATIONS (1983) (made pursuant to 49 U.S.C. § 2223) (1982).

system which is typically supplied by one large carrier. It has been charged that if a carrier denies access to a competitor, uses market information obtained from the reservation system to pressure travel agents to ticket on its own airline, or adopts other anticompetitive practices, the beneficial effects of entry deregulation can be stymied, at least in the short run.[46]

Second, contestability theory has important implications for merger policy. Any merger involving overlapping routes would never have been approved under the 1968 Justice Department guidelines, since, as we have mentioned, four-firm concentration ratios for virtually all city-pair routes are near one hundred percent.[47] Nevertheless, ease of entry and exit ought to be a uniform characteristic of airline markets. Thus, the assessment of mergers for such markets should rely on a functional analysis of the degree of contestability of markets rather than on market share and concentration ratio data. As long as there are comparable airlines with stations at one or both ends of the overlapping markets, the CAB does not consider a competitive problem to be present. Bailey[48] cites the case of the Houston-New Orleans market in the Texas International and National merger case. In spite of the two-firm concentration ratio of seventy-five percent after the proposed merger, the presence of eleven carriers with facilities already functioning at both ends of this market led to approval of the merger. Contestability theory indicates that it is precisely that sort of case-by-case analysis, taking into account ease of entry and exit as well as scale effects within markets, that should be used by policy-makers to evaluate the appropriateness of mergers. Indeed, it appears that the Justice Department has moved in this direction.[49]

46. In December 1982, the Conference Committee on the 1983 appropriations for the Department of Transportation was sufficiently concerned about this issue that it ordered a joint CAB and Department of Justice investigation into computer reservation system practices. *See* 128 CONG. REC. H9510, H9515 (daily ed. Dec. 13, 1982) (conference report on H.R. 7019); CIVIL AERONAUTICS BOARD, REPORT TO CONGRESS ON AIRLINE COMPUTER RESERVATIONS SYSTEMS (1983).

47. *See* Merger Guidelines of Department of Justice, 2 TRADE REG. REP. (CCH) ¶4510, at 6884 (May 30, 1968). However, the 1982 Merger Guidelines, 47 Fed. Reg. 28,493, explicitly consider potential entry both in identifying the firms to be included in the relevant market, 47 Fed. Reg. 28,495, and in assessing the ability of existing firms to raise price, 47 Fed. Reg. 28,498. Consideration of the prospects for entry into airline markets might well lead the Justice Department not to challenge an airline merger. This, of course, is entirely consistent with the implications of contestability analysis.

48. Bailey, *Contestability and the Design of Regulatory and Antitrust Policy*, 71 AM. ECON. REV. 178, 181 (1981). For a fuller description, see CIVIL AERONAUTICS BOARD, ANTITRUST POLICY FOR THE AVIATION INDUSTRY (1982).

49. 1982 Merger Guidelines, *supra* note 47, at 28,495-98.

C. *Trucking Regulation and its Partial Reform*

Theory suggests that, with the possible exception of barge transportation, trucking should be perhaps the most contestable of the economy's industries; but it has long been subjected to regulatory entry control. The largest class of ICC-regulated carriers—general freight carriers—specialize in less-than-truckload services, making multiple deliveries and pick-ups along regular routes. Regulatory control has taken the form of certificates describing the commodities permitted to be hauled and the specific cities along which each such commodity may be carried. These interventions have undercut the ability of carriers to serve routes at minimum cost since the restrictions have made it difficult to utilize capacity fully, especially on backhauls. Breyer cites studies indicating that regulated general-freight vans return empty more than one-third of the time, and that a reduction in their regulation could increase load factors by ten percent.[50] Studies also indicate that ICC restrictions on the ability of "exempt" carriers (such as private carriage by firms of their own goods, carriage of goods within states and carriage of agricultural commodities) to haul goods for movements subject to regulation has resulted in an excessive amount of empty backhauling.[51]

The Motor Carrier Act of 1980[52] liberalized entry into the less-than-truckload portion of trucking and exempted additional types of motor carrier transportation from economic regulation.[53] From the perspective of contestability, these route liberalizations offer an important opportunity for enhanced efficiency. In a recent study that attempts to distinguish between economies of scale and economies of scope in the trucking industry, Chiang offers important new evidence about how greater freedom of entry may contribute to market efficiency.[54] She found that there are strong economies of joint production in distribution networks associated with short-haul and intermediate-haul trucking shipments, particularly for small and medium-sized firms. Mergers in trucking are undertaken by firms seeking to obtain the full range of benefits afforded by these economies of scope. These findings illustrate an important feature of contestable markets: if left alone, contestable markets will tend to move toward the

50. *See* S. BREYER, *supra* note 36, at 225 nn. 47-48.

51. *See* P. MACAVOY & J. SNOW, REGULATION OF ENTRY AND PRICING IN TRUCK TRANSPORTATION 24-27 (1977).

52. Pub. L. No. 96-296, 94 Stat. 793 (codified as amended in scattered sections of 49 U.S.C.).

53. Additional exemptions are set forth in 49 U.S.C. § 10,526(a). Entry restrictions are eased by *id.* §§ 10,922, 10,762, & 11,145.

54. W. Chiang, Economies of Scale and Scope in Multiproduct Industries: A Case Study of the Regulated U.S. Trucking Industry (1981) (Ph.D. Dissertation, Dep't of Civil Engineering, MIT). For a discussion of other cost studies of multiproduct industries, see Bailey & Friedlaender, *Market Structure and Multiproduct Industries*, 20 J. ECON. LITERATURE 1024 (1982).

most efficient organization of productive forces.

Chiang's study shows also that there is no evidence of global economies of scale in trucking, which suggests that such firms will not attain monopoly size in an unregulated environment. Thus, regulatory inhibition of entry in trucking has been perverse in at least two respects: it has promoted inefficiency in market structure, and it has been unnecessary for the control of monopoly. Moreover, there is no evidence that barriers to entry would have emerged in trucking had regulatory intervention not occurred. Most of the trucking industry's costs are variable, consisting of trucks and drivers. The consolidation of shipments by networking or centralizing repair or administration can produce economies of scale, but these are unlikely to confer monopoly power; firms benefitting from these economies cannot raise prices much above their costs without losing business to enterprises on nearby routes that could readily extend their operations.

The Motor Carrier Act of 1980 introduced zones of price flexibility for general commodity carriage, thereby somewhat increasing the opportunity for competition.[55] Available evidence suggests that decontrol would lead to lower prices. For example, when poultry and frozen foods and vegetables were reclassified in the 1950's to fall under an agricultural exemption, rates declined over a five-year period by thirty-three percent for poultry and nineteen percent for frozen foods.[56] Breyer cites studies that show that rates are lower in intrastate than in interstate markets,[57] and that rate reductions followed the relaxation of regulations in countries such as West Germany and Australia.[58]

D. *Telecommunications Regulation and the Antitrust Case*

Although Congress has now considered legislative reform of telecommunications regulation for several years, the most dramatic regulatory changes have actually emerged from the Federal Communications Commission (FCC), the Department of Justice and the federal courts. The decisions can be characterized as attempts to isolate the segment of the telecommunications market still considered to involve technological natural monopoly, i.e., local telephone service, from segments that can, perhaps, no longer be taken to constitute natural monopolies, such as long distance services and the provision of terminal equipment. The situation has been

55. 49 U.S.C. § 10,708.

56. J. SNITZLER & R. BYRNE, INTERSTATE TRUCKING OF FRESH AND FROZEN POULTRY UNDER THE AGRICULTURAL EXEMPTION (1958) (Dep't of Agric. Mktg. Research Div., MRR-244); J. SNITZLER & R. BYRNE, INTERSTATE TRUCKING OF FROZEN FRUITS AND VEGETABLES UNDER THE AGRICULTURAL EXEMPTION (1959) (Dep't of Agric. Mktg. Research Div., MRR-316).

57. *See* S. BREYER, *supra* note 36, at 229 n.52.

58. *Id.* at 229 nn.54-55.

complicated primarily by the large costs common to local and long distance services and by the possible economies and externality benefits associated with joint operation of these services within a single firm.

FCC and court decisions that permit subscribers to buy terminal equipment from firms other than AT&T's subsidiary, Western Electric, and to attach the equipment to AT&T's lines raise few problems in terms of contestability analysis.[59] So long as production of this equipment is not a natural monopoly and the equipment manufactured by firms other than Western Electric does not produce noise or other forms of "harm" for the telephone network, contestability analysis suggests that regulatory policy should encourage access to local telephone systems on equal terms because of the large sunk costs associated with the provision of local service. Freedom of entry is preferred, according to the theory, over arrangements that bar local phone companies or other firms from selling telephone equipment.

The issues involving local versus long distance services are more complex, in part because of the common cost problem and in part because of the efficiencies derived from coordinated operation of an integrated network. For example, such an efficiency arises because AT&T commonly routes calls during busy periods through distant switching centers if nearer ones are operating at full capacity.[60] This is only one of a variety of network-wide planning decisions that may make production less costly when local and long distance operations are contained within one firm.[61]

On the other hand, the natural monopoly characteristics of long distance services and the large sunk costs associated with those services are being modified by technological change. Cable technology has at least to some extent been replaced in long distance markets by wireless microwave transmissions systems and, more recently, by satellites. These new techniques have opened new policy options, since, for example, it may be possible for several different firms to use microwave or satellite technology to transmit calls on many routes without significantly increasing unit costs. In a series of decisions beginning in 1959 with *Above 890*, the FCC has given a number of firms the right to use microwave transmissions for private line services, i.e., services not involving any connection with the Bell local exchange network.[62] The D.C. Circuit, in the 1977-78 *Execunet*

59. Hush-a-Phone Corp. v. United States, 238 F.2d 266 (D.C. Cir. 1956); Carterphone Device, 13 F.C.C.2d 420, 14 F.C.C.2d 571 (1968); Interstate Foreign Message Toll Tel. Serv. (Registration Program), 56 F.C.C.2d 593 (1975), 58 F.C.C.2d 736 (1976).

60. BELL TELEPHONE LABORATORIES, INC., ENGINEERING AND OPERATIONS IN THE BELL SYSTEM 31, 87-92 (1977).

61. *See* Southern Pac. Com. Co. v. AT&T Co., 556 F. Supp. 825, 868-70 (D.D.C. 1983).

62. Allocation of Frequencies in the Bands Above 890 Mc., 27 F.C.C. 359 (1959), *modified*, 29 F.C.C. 825 (1960). *See, e.g.*, Microwave Com., Inc., 18 F.C.C.2d 953 (1967), *reconsid. denied*, 21

rulings, overturned FCC decisions and extended freedom of entry beyond the provision of private line services, permitting entry in direct competition with Bell long distance service.[63] The *Domestic Satellites* decision permitted firms other than AT&T to use satellites to provide specialized communications services.[64] Similarly, emerging technological developments such as cellular mobile radio are threatening to erode the natural monopoly attribute of local service.[65]

The consent decree involving the Justice Department and AT&T[66] is far more wide-ranging than anything done by the FCC, as it imposed a major change in market structure. It required a complete separation of local and long distance services. Local service is now offered by a series of regulated monopolies, consisting of groupings of former Bell System operating companies. Long distance service is provided in a much more competitive environment. The Bell System is also permitted to enter unregulated markets, such as the computer market, in which the technology is becoming less and less distinguishable from that in communications.

Contestability analysis can provide some framework for discussion of the issues surrounding local and long distance service, but it does not necessarily offer definitive answers. To the extent that the consent decree has succeeded in separating markets characterized by sunk costs and natural monopoly from markets that are reasonably contestable, the decree would seem consistent with contestability analysis. Similarly, contestability analysis favors the realization of economies of scope between telecommunication and computer services which the decree permits. However, to the extent the decree prevents the realization of economies of scope between local and long distance services or results in substantial quality differences across regions with attendant degradation in all service or prevents competition when new technologies, such as cellular mobile radio, reduce sunk costs requirements, questions are raised by the analysis. These uncertainties reflect the critical importance of technological considerations in the application of contestability theory.

F.C.C.2d 190 (1970), *modifs. granted,* 27 F.C.C.2d 380 (1971); Specialized Common Carrier Servs., 29 F.C.C.2d 870 (1971), *aff'd sub nom.* Washington Util. & Transp. Comm'n v. FCC, 513 F.2d 1142 (9th Cir.), *cert. denied,* 423 U.S. 836 (1975).

63. MCI Telecom. Corp. v. FCC (*Execunet I*), 561 F.2d 365 (D.C. Cir. 1977), *cert. denied,* 434 U.S. 1040 (1978); MCI Telecom. Corp. v. FCC (*Execunet II*), 580 F.2d 590 (D.C. Cir.), *cert. denied,* 439 U.S. 980 (1978).

64. Domestic Communications-Satellite Facils., 35 F.C.C.2d 844, 38 F.C.C.2d 665 (1972). *See also* United States v. FCC, 652 F.2d 72 (D.C. Cir. 1981).

65. Allocation of Frequencies in the 150.8-162 Mc/s Band, 12 F.C.C.2d 841, 14 F.C.C.2d 269 (1968), *aff'd sub nom.* Radio Relay Corp. v. FCC, 409 F.2d 322 (2d Cir. 1969); Mobile Radio Communications, Inc., 29 F.C.C.2d 62 (1971); Cellular Communications Sys., 50 RAD. REG. 2D (P & F) 1673, 51 RAD. REG. 2D (P & F) 143 (1982).

66. United States v. AT&T Co., 552 F. Supp. 131 (D.D.C. 1982), *appeal dismissed,* United States v. Western Elec. Co., 714 F.2d 178 (D.C. Cir. 1983).

Conclusion

The new contestability theory clearly makes no pretense of solving all problems, but it does seek to identify the proper questions—how to identify circumstances in which deregulation should occur and, where continuing regulation is appropriate, what forms it should take. By focusing attention on sunk costs as a major reason for regulatory intervention, and by specifying a variety of tools and methods to minimize the market power associated with them, the contestability perspective offers some degree of direction to policymakers. Policy analysts should begin by determining what regulatory or other obstacles stand in the way of contestability and then should consider ways to reduce or eliminate them.

Contestability theory also offers a coherent analysis of market structure issues, and underscores how important investigation of actual conduct and performance is in the presence of concentration attributable to scale economies. Because there may well be sunk costs and entry barriers at an industry level, a market-by-market analysis within the industry may have to be undertaken. The new theory supports policy measures that attempt to separate out those portions of an industry in which market failures attributable to natural monopoly or other elements play an important role from those portions of the industry in which fixed, but not sunk, costs predominate, so that competition and consumer choice can contribute to quality and restrain costs. If particular markets are readily contested, there may be no need for continued intervention in these markets. Similarly, the theory suggests that even in industries where deregulation may be called for in some areas, continued regulation to maintain open access may be appropriate. In this and other ways the new theory sheds light on the appropriateness of traditional patterns of regulation and the avenues available for deregulation.

Our discussion of particular regulated industries confirms once again that reality is more complex than any theoretical model, so that the latter can never be expected to provide standardized procedures that produce cut-and-dried solutions to the problems encountered in practice. We may seek to determine appropriate boundaries between regulated and unregulated portions of an industry, but boundaries based on technological considerations alone may prove misplaced. An analysis such as that provided by contestability theory, while not free of difficulties, can reasonably aspire to offer the practitioner greater confidence and clearer insight into the pertinent issues.